REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	August 1999	Final Report 29/Sept/95 to July 31, 1999

4. TITLE AND SUBTITLE
Some Topics in Data Analysis and Stochastic Modeling

5. FUNDING NUMBERS G
DAAH04-95-1-0665

6. AUTHOR(S)
James R. Thompson

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rice University, Department of Statistics
Houston, Texas 77005

8. PERFORMING ORGANIZATION REPORT NUMBER
R124000

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U. S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 34522.1-MA

11. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision, unless so designated by other documentation.

12 a. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12 b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

The advent of highspeed computing on the PC provides new possibilites for data based modeling. One key is simulation. Simulation gives us a venue for dealing with parameter estimation in stochastic models not previously tractable. We consider examples from oncology, economics, statistical process control and epidemiology. We essentially consider two realities, the first consisting of random nonparametric interpolations of the data and the second random implementations based on a mathematical model. We use simulation to change the model parameters to bring the reality of the model to consistency with that of the data. Also, high speed computing enables us to carry out nonparametric data analysis in higher dimensions. The key here is to build algebraic rather than geometrical algorithms. We show how nonparametric data analysis in high dimensions (say greater than three) should generally not be treated as one of nonparametric density estimation but rather should start with finding centers of relatively high density and then use parametric approximations in the neighborhoods of these modes.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Simulation, Modeling, Oncology, Economics, Epidemioiogy	4
	16. PRICE CODE

17. SECURITY CLASSIFICATION OR REPORT	18. SECURITY CLASSIFICATION ON THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

NSN 7540-01-280-5500

Standard Form 298 (Rev.2-89)
Prescribed by ANSI Std. 239-18
298-102

19991103 116

Enclosure 1

DTIC QUALITY INSPECTED 4

Final Report (ARO-DAAH04-95-1-0665) *Some Topics in Data Analysis and Stochastic Modeling*

(4) Statement of the problem studied.

As stated in the Abstract of the original proposal: "The analysis of data absent some consideration for the models of the systems which generated them frequently causes the analysis to be less effective than it could be. Much of the work in nonparametric function estimation has been content to stop at the level, say, of a nonparametric regression curve which gives little clue as to the underlying processes. We propose to continue a line of research in which model building and data analysis proceed hand in hand."

The purpose of this research has been to develop methodologies where models are developed and tested in the light of data. We have used a number of settings to develop this work.

(5) Summary of the most important results.

The book *Simulation: A Modeler's Approach* [9] embodies a general theory for model based simulation.

We have proposed a Bayesian-Paretan quality optimization model for use in NASA's Space Station project [5]. This is a prospective model, as present, since the Space Station is yet to be constructed. We have also developed models for dealing with quality improvement via statistical process control in situations where the output data to be examined is of high dimensionality [1–2].

We have also noted discrepancies between the efficient market theory models of neoclassical economics and the reality of the market [7, 9, 13, 14]. Management Science frequently overlooks the fact that the market is based on the production of real goods and services, which production should be constantly under improvement. We have developed an integrated approach in which entrepreneurship, finance and statistical process control are employed, [7, 9, 13, 14]. We have worked and are working, for example, to improve the efficient market model of Black, Scholes and Merton—which model has produced poor results, for example, the failure of LCTM. The modeling of economic phenomena is one of those which is made much more tractable by simulation based techniques. We have developed [3] forecasting models for natural gas prices which outperform the algorithms currently in use..

We have used the AIDS data base of the WHO AIDS Working Group to determine the driving force of AIDS in the First World. Our conclusion, based on the data, is that it is the massive United States AIDS infective pool which, due to the relative cheapness of international travel, drives the AIDS epidemic in Canada and Europe [6, 8, 9]. Also, we examined the danger of the immunosuppression of HIV in facilitating an adjuvant tuberculosis epidemic [10–12].

We have shown how the problem of nonparametric data analysis with higher dimensional data can be structured by first finding centers of high density and then using Gaussian approximations locally. We have observed that the "curse of dimensionality" is frequently not real when one uses the mode-finding preprocessing step. We have used mode finding approaches to investigate Army ballistics data of high dimensionality and found that one can perform very well in, say, eight space with samples in the few hundreds range [9, Chapter 11].

We have demonstrated how the SIMEST [4, 9] algorithm of Thompson can be used in practical situations to estimate the parameters of deep models using simulation-based techniques.

(6) List of Publications published under (ARO-DAAH04-95-1-0665.

1. Koronacki, J. and Thompson, J.R. (1996). "Statistical process control viewed as a stepwise procedure for the improvement of quality in the production and service industries" in *Problemy Jakości*, v. xxvii, pp. 6-12.

2. Koronacki, J. and Thompson, J.R. (1997). "Statistical process control for processes with multiple-dimensional measurements," in *Problemy Jakości*, v. xxix, pp. 8-15.

3. Lawera, M. (1999). *A Stochastic Model for Pricing Natural Gas Contracts*, doctoral dissertation, Houston: Rice University.

4. Schwalb, O. (1999). *Practical and Effective Methods of Simulation Based Parameter Estimation for Multidimensional Data*, doctoral dissertation, Houston: Rice University.

5. Thompson, J.R. and Walsh, R. (1996)."A Bayesian Pareto analysis for system optimization" in *Proceedings of the First Annual U.S. Army Conference on Applied Statistics*, B. Bodt., ed. Research Triangle, N.C.: Army Research Office, pp. 71-83.

6. Thompson, J.R., (1998). "The United States AIDS Epidemic in First World Context," (1998) in *Advances in Mathematical Population Dynamics: Molecules, Cells and Man*, O. Arino, D. Axelrod and M. Kimmel, eds. Singapore: World Scientific Publishing Company, pp. 345-354.

7. Thompson, J.R. and Williams, E.E. (1999). "A Post Keynesian analysis of the Black-Scholes option pricing model" in *The Journal of Post Keynesian Economics* (in press).

8. Thompson, J.R. (1999). "Understanding the AIDS Epidemic: A Modeler's Journey," in *Applied Mathematical Modeling*, T. Wallenius and D. Shier, eds. CRC Press (in press).

9. Thompson, J.R. (1999). *Simulation: A Modeler's Approach*, New York: John Wiley & Sons, 1999, 297 pages (in press).

10. West, R.W. and Thompson, J.R. (1998). "Modeling the impact of HIV on the spread of tuberculosis in the United States," in *Mathematical Biosciences* 143:35-60.

11. West, R.W. and Thompson, J.R. (1998). "Models for the Simple Epidemic" in *Mathematical Biosciences* 141:29-39.

12. West, R.W. and Thompson, J.R. (1998). "Model based back calculation for the HIV epidemic" in *Advances in Mathematical Population Dynamics: Molecules, Cells and Man*, O. Arino, D. Axelrod and M. Kimmel, eds. Singapore: World Scientific Publishing Company, pp. 329-343.

13. Williams, E.E. and Thompson, J.R. (1996). *The Economics of Production and Productivity: A Modeling Approach*, Austin, Texas: Capital, 443 pages.

14. Williams, E.E. and Thompson, J.R. (1998). *Entrepreneurship, Planning, and Productivity*, Philadelphia, Pennsylvania: University Press of America, 1998, 342 pages.

(7) List of participating personnel showing any advanced degrees earned by them while employed on the project.

Martin Lawera received support in 1995. His doctoral dissertation (*A Stochastic Model for Pricing Natural Gas Contracts*), Houston: Rice University.) was accepted in summer of 1999

Otto Schwalb, III received support in 1996-1999. His doctoral dissertation (*Practical and Effective Methods of Simulation Based Parameter Estimation for Multidimensional Data*) was accepted in summer of 1999.

James R. Thompson, principal investigator received an average of 1.5 months of summer support over the period of the grant.

(9) Bibliography (please see (6).)